LLM

List of Curated Research Papers

Arockia Liborious

BROAD *Category*

01 Model Architecture and Efficiency

02 Preference Optimization and RLHF

03 Multimodal Models

04 Long-Context Learning and Inference

05 Reasoning and Knowledge

06 Quantization and Compression

07 Evaluation and Benchmarks

08 Instruction Tuning and Alignment

09 Surveys and Meta-Analysis

10 Applications and Tooling

CLICK HERE

Introduction

This curated list highlights groundbreaking advancements in large language models (LLMs) across ten categories, driving progress toward AGI. From efficient architectures to multimodal reasoning and robust evaluation methods, these papers represent pivotal steps in aligning, scaling, and optimizing LLMs for general intelligence. Each selection showcases innovative approaches addressing real-world challenges, from improving reasoning to democratizing access to cutting-edge tools. These works collectively shape the trajectory of AGI development, fostering collaboration and pushing the boundaries of what intelligent systems can achieve.

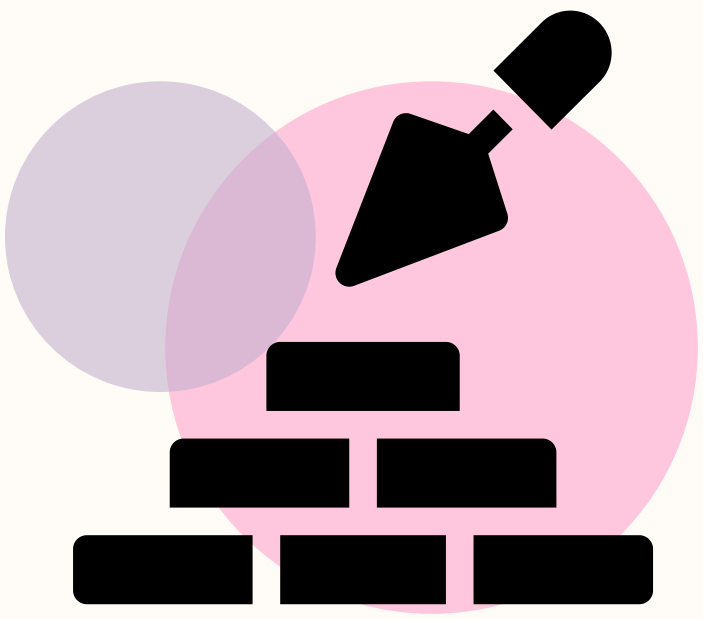
The Purpose Of This Content

Highlighting transformative papers, my personal favorites, showcasing the future potential of AGI through groundbreaking LLM advancements.

01

MODEL ARCHITECTURE AND EFFICIENCY

PAPER



Mixture-of-Transformers: A Sparse and Scalable Architecture for Multi-Modal Foundation Models

CLICK HERE

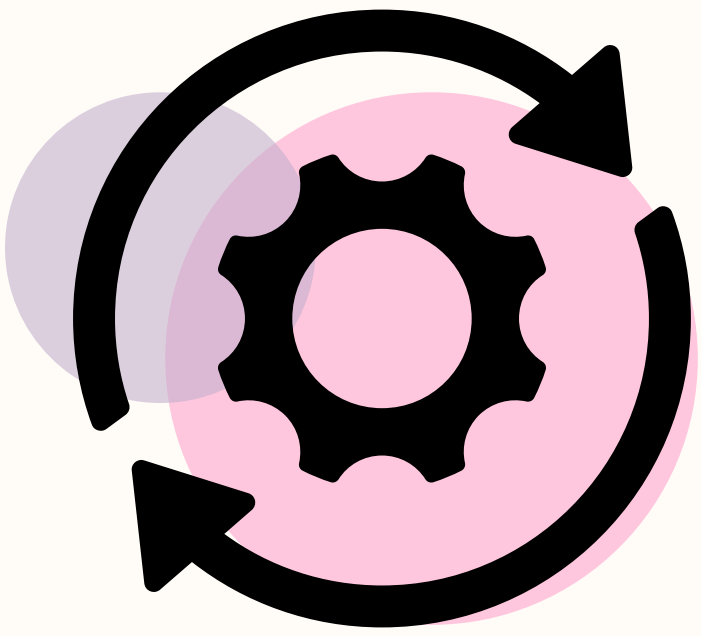
WHY

Introduces a sparse transformer architecture that scales efficiently across multiple modalities, significantly reducing computational costs while enabling AGI-like multitasking capabilities.

02

PREFERENCE OPTIMIZATION AND RLHF

PAPER



The Perfect Blend: Redefining RLHF with Mixture of Judges

CLICK HERE

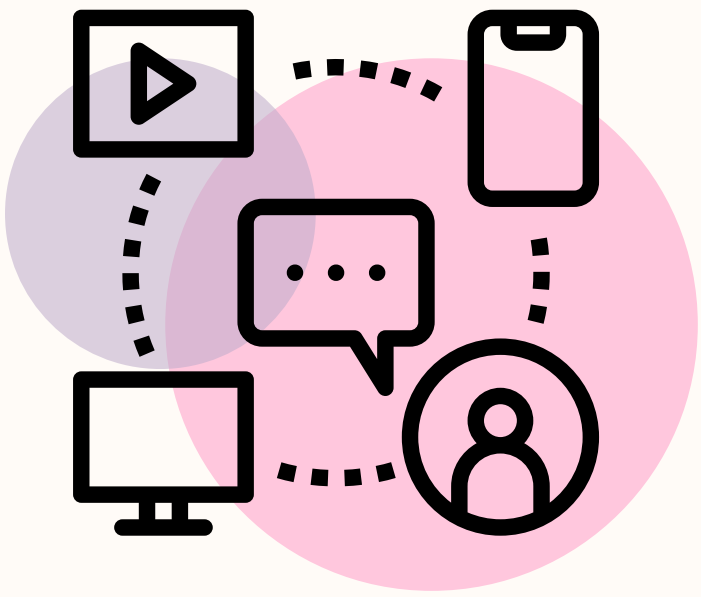
WHY

Redefines reinforcement learning with human feedback by incorporating multiple feedback sources, creating systems that can better understand and align with nuanced human values, a critical step toward AGI safety and alignment.

03

MULTIMODAL MODELS

PAPER



ARIA: An Open Multimodal Native Mixture-of-Experts Model

CLICK HERE

WHY

By natively combining multiple modalities into a single framework, this paper pushes LLMs closer to AGI by enabling them to integrate and reason across diverse data types like images, text, and audio.

04

LONG-CONTEXT LEARNING AND INFERENCE

PAPER

RetrievalAttention: Accelerating Long-Context LLM Inference via Vector Retrieval

CLICK HERE

WHY

Tackles the challenge of long-context inference by introducing a novel attention mechanism, enabling models to process vast amounts of contextual information—a cornerstone for AGI-level reasoning.

05

REASONING AND KNOWLEDGE

PAPER

Towards Large Reasoning Models: A Survey of Reinforced Reasoning with Large Language Models

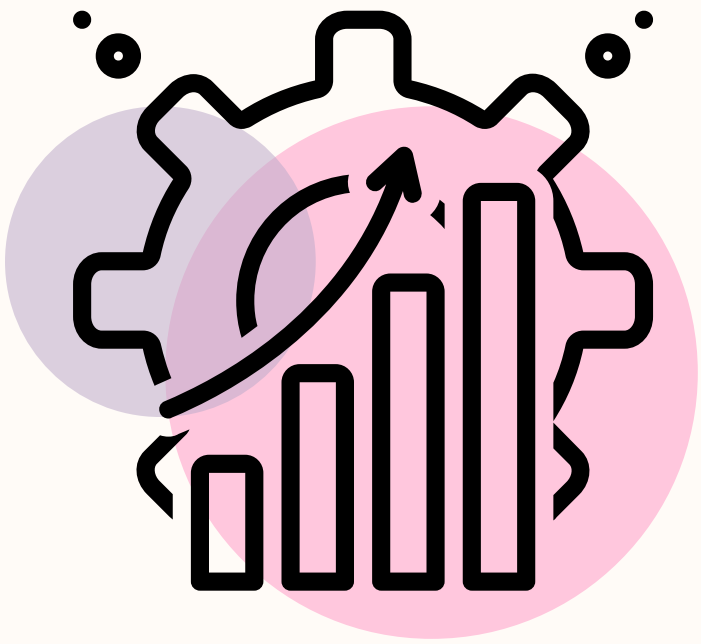
CLICK HERE

WHY

Offers a comprehensive framework for enhanced reasoning in LLMs using reinforcement learning, paving the way for AGI systems capable of solving complex, real-world problems autonomously.

06

QUANTIZATION AND COMPRESSION



PAPER

BitNet a4.8: 4-bit Activations for 1-bit LLMs

CLICK HERE

WHY

Achieves groundbreaking efficiency in LLM deployment without sacrificing performance, making AGI systems more scalable and accessible.

07

EVALUATION AND BENCHMARKS

PAPER

Adding Error Bars to Evals: A Statistical Approach to Language Model Evaluations

CLICK HERE

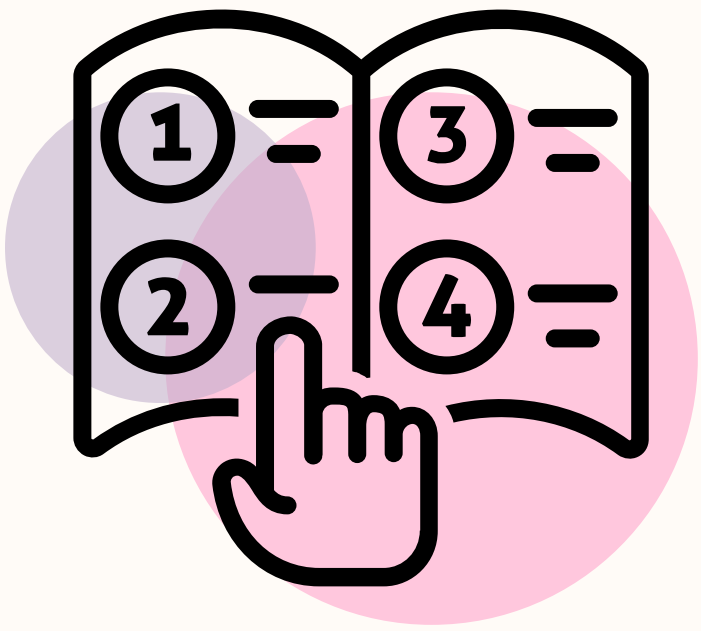
WHY

Introduces robust statistical methods for evaluating LLMs, ensuring AGI systems are reliably assessed for their capabilities, safety, and alignment with human goals.

08

INSTRUCTION TUNING AND ALIGNMENT

PAPER



Instruction Following without Instruction Tuning

CLICK HERE

WHY

Demonstrates that LLMs can achieve instruction-following capabilities without explicit tuning, simplifying the path to AGI systems that can generalize across tasks seamlessly.

09

SURVEYS AND META-ANALYSIS

PAPER

A Comprehensive Survey of Small Language Models in the Era of Large Language Models

CLICK HERE

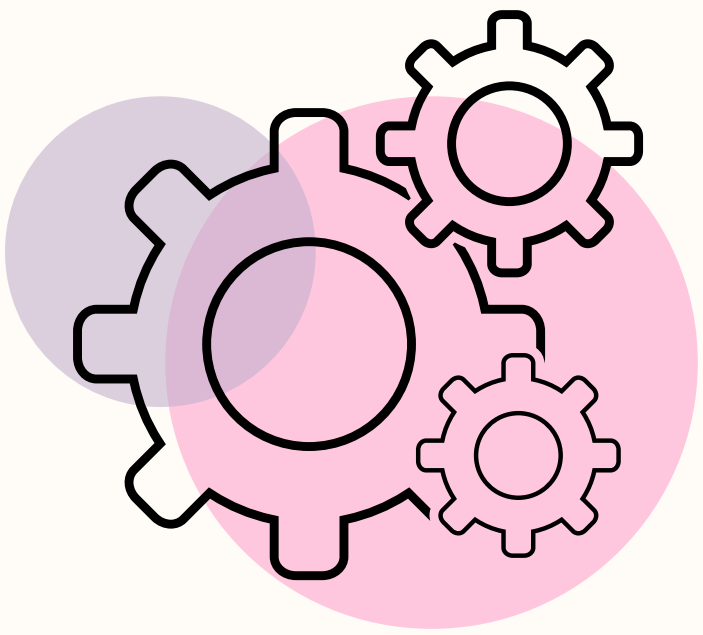
WHY

Highlights the complementary role of smaller models in AGI research, showcasing how hybrid systems (small and large models) can bridge the gap toward achieving general intelligence.

10

APPLICATIONS AND TOOLING

PAPER



RedPajama: An Open Dataset for Training Large Language Models

CLICK HERE

WHY

Democratizes access to high-quality training datasets, enabling researchers worldwide to contribute to AGI development, fostering a more inclusive and collaborative AI ecosystem.

BONUS

PAPER

LLaMA: Open and Efficient Foundation Language Model

LLaMA (Large Language Model Meta AI) is a collection of foundation language models introduced by Meta AI

CLICK HERE

- **Performance:** LLaMA-13B outperforms GPT-3 (175B); LLaMA-65B rivals Chinchilla-70B and PaLM-540B.
- **Accessibility:** Open-source release fosters innovation and democratizes LLM access.
- **Efficiency:** Smaller models trained on more data achieve better performance, reducing costs and improving scalability.

FOR MORE SUCH

CONTENTS

Follow me